

ABSTRACT

An improved system and method for building a large index is provided. The system and method may be used by many
5 applications to build a large index, including a search engine for crawling the World Wide Web. An indexing engine with an index merger may build an index of content by using a staged pipeline for merging sub-indexes. The index merger may concurrently merge sub-indexes created at multiple stages
10 during indexing of content by using threads from a merging thread pool. When all the content has been indexed, the system may proceed to perform a final merge of all available sub-indexes to form a master index. The system and method may build a large index of any type of content including
15 documents, images, audio streams, video streams and other types of content.